



Investigación en
Educación Médica

<http://riem.facmed.unam.mx>



EDITORIAL

«Mi instrumento es más válido que el tuyo»: ¿Por qué seguimos usando ideas obsoletas?



“My instrument is more valid than yours”: Why do we continue using obsolete ideas?

«Es incorrecto usar la frase “la validez del examen”...»

Standards for Educational and Psychological Testing.
AERA, APA, NCME. 2014, p. 23.

«Nada hay más fuerte que el hábito».

Ovidio

Uno de los conceptos más importantes en evaluación educativa es el de validez (para algunos autores, es el más importante), ya que permea por todos lados el proceso educativo y determina la congruencia interpretativa del uso de los resultados de pruebas y exámenes de acuerdo a los fines para los que fueron diseñados. A pesar de ello, muchos colegas docentes en las profesiones de la salud no están familiarizados con los aspectos modernos del concepto de validez y las condiciones del «estado del arte» sobre el tema. ¿Cuántas veces ha escuchado el lector variantes de las siguientes frases, proferidas con asertividad por funcionarios educativos, directores, maestros, investigadores y académicos de diversas disciplinas?

- «Utilizamos un instrumento validado, por lo que los resultados son confiables».
- «El examen que aplicamos ya fue validado por expertos».
- «Nuestro examen es válido porque tiene una alta confiabilidad».
- «El alfa de Cronbach es alta, así que los resultados son válidos».
- «Ya validamos nuestro cuestionario».

Todas estas oraciones implican un uso inapropiado del concepto moderno de validez, por lo que sería deseable que las dejáramos de usar (tarea nada sencilla en nuestro medio,

La revisión por pares es responsabilidad de la Universidad Nacional Autónoma de México.

como le consta al autor de estas líneas). Los seres humanos nos acostumbramos a utilizar frases que con la fuerza del uso se convierten en rutina, en costumbre sancionada por el paso del tiempo. Si a ello agregamos que estas frases son expresadas por personas con algún grado académico, con prestigio y poder en la estructura educativa de que se trate, dichas ideas son absorbidas por los estudiantes y el grupo de trabajo, convirtiéndose así en dogma indiscutible y repitiéndose de manera instintiva e irreflexiva. Es fascinante cómo un concepto tan importante y central para la educación como es la validez puede ser malentendido de esta manera, sobre todo porque en las últimas 2 décadas se han publicado a nivel mundial múltiples actualizaciones importantes sobre el tema^{1,2}.

En la primera mitad del siglo pasado se gestó el concepto de validez, reconociéndose principalmente la validez de contenido y la de criterio³. Posteriormente se cayó en la cuenta de que la validez de contenido prácticamente siempre apoyaba a la prueba, y que identificar un estándar de referencia para atributos intangibles no era sencillo, por lo que fue propuesta la validez de constructo. Con esto se conformó la «trinidad clásica» de la validez: de contenido, de criterio y de constructo, que ha sido la terminología con la que la mayoría de los psicólogos, pedagogos y profesionales de la medición se formaron el siglo pasado. Este modelo conceptual se convirtió en «lugar común» en el lenguaje cotidiano de la evaluación y medición educativa.

El esquema mencionado ha sido utilizado por varias décadas en la comunidad de educadores, pero tiene una serie de inconvenientes: genera una separación artificial entre los «diferentes tipos» de validez, dando la impresión que son cosas diferentes e independientes entre sí; propicia el concepto erróneo de que los exámenes y las pruebas son válidos o inválidos por sí mismos, implicando que la validez es una propiedad intrínseca del instrumento que

<http://dx.doi.org/10.1016/j.riem.2016.05.007>

2007-5057/Derechos Reservados © 2016 Universidad Nacional Autónoma de México, Facultad de Medicina. Este es un artículo de acceso abierto distribuido bajo los términos de la Licencia Creative Commons CC BY-NC-ND 4.0.

podría ser transferible a otros contextos; se asigna a la confiabilidad un papel más importante del que realmente tiene, incluso separándola de la validez. A finales de los 80, Messick propuso abandonar el modelo de los diferentes «tipos» de validez, para migrar a un marco conceptual unificado en el que toda la validez es validez de constructo, que se alimenta de diferentes fuentes (contenido del examen, proceso de respuesta, estructura interna, relación con otras variables, consecuencias del examen)². Este autor afirmó: «Validez es un juicio evaluativo e integrado del grado al que la evidencia empírica y razonamientos teóricos apoyan lo adecuado y apropiado de las inferencias y acciones basadas en puntuaciones de exámenes y otras formas de evaluación»². Este modelo holístico de validez ha sido ampliamente aceptado por la comunidad internacional de investigadores en evaluación, al grado que en la edición de 1999 de los *Standards for Educational and Psychological Testing*, el documento más importante sobre este tema a nivel mundial, fue incluido explícitamente en sus lineamientos¹.

El modelo más reciente de validez, basado en el propuesto por Messick, ha sido desarrollado por Michael Kane en EE. UU.⁴. Kane enfatiza que la validez se refiere a las inferencias que se hacen al usar los resultados de la aplicación de exámenes para fines específicos, y que la validez no es una propiedad intrínseca del instrumento, examen o prueba. En sus palabras: «Validar una interpretación propuesta o un uso de los puntajes de exámenes es evaluar las afirmaciones basadas en esas puntuaciones. La mezcla específica de evidencia necesaria para la validación depende de las inferencias que se hagan y los supuestos que se utilicen»⁴, de tal manera de que un examen no es válido o inválido *per se*, lo que es más o menos válido es la fortaleza de las inferencias que se hacen del uso de los resultados de las pruebas. Es importante resaltar que los usuarios de los resultados de los exámenes deben tener claro para qué los van a usar, y utilizar la lógica y la metodología científica para documentar lo apropiado del uso específico que se haga de esos resultados. Se recomienda al lector consultar las excelentes revisiones de Downing, Cook et al. y Kane sobre el tema²⁻⁴.

Algunos de los beneficios del uso del concepto moderno de validez en educación son: enfoca la atención en una amplia gama de temas asociados con la *interpretación* y el *uso* de los resultados de exámenes; enfatiza que nos basamos en una serie de *premisas* cuando interpretamos resultados, y que necesitamos *chechar* nuestras premisas; además permite interpretaciones y usos alternos de los resultados de exámenes. Por lo tanto no es un asunto de mera semántica o filigrana verbal intrascendente, la comprensión cabal de estos conceptos es necesaria para el diseño apropiado de los instrumentos de evaluación y el uso adecuado de sus resultados. La forma en que nos expresamos al usar estos términos tiene un impacto importante en la manera como aplicamos e interpretamos los conceptos de evaluación en la práctica.

Desafortunadamente estos avances conceptuales han penetrado con lentitud en la comunidad de educadores de profesiones de la salud a nivel global. Esto puede ser por múltiples razones, como el uso del idioma inglés en las publicaciones originales (que de por sí no son fáciles de leer y en ocasiones están cargadas de matemáticas avanzadas), o el

hecho de que la mayoría de los profesionales de la salud no estamos acostumbrados a consultar la literatura científica original de medición educativa y evaluación en nuestra práctica docente (además de que muchas instituciones de salud no tienen acceso a este tipo de publicaciones en sus bibliotecas).

¿Por qué toda esta perorata?, porque a pesar de que el concepto de validez ha evolucionado de manera importante en el «mundo cuantitativo», en el que todavía tenemos un largo trecho por andar, en la investigación cualitativa el concepto tiene aristas de diversa índole todavía más complejas. En este número de la revista, las maestras Tania Vives y Margarita Varela realizan un fascinante análisis sobre lo complicado de aplicar los conceptos tradicionales de validez en los estudios cualitativos, con algunas propuestas interesantes. El tema es de enorme vigencia, ya que recientemente en el contexto de la educación médica, Cook et al. han analizado con creatividad esta temática^{3,5}.

Como corolario, sugiero a nuestros amables lectores que cuando escuchen a algún colega hablar de «un instrumento válido», lo inviten (amablemente y con la prudencia que el caso requiera) a actualizarse y sumergirse en la interesante literatura sobre el tema. Al momento de escribir estas líneas, identifiqué algunos artículos publicados en nuestra revista que cometen el error conceptual arriba descrito!

Este número de la revista incluye además un fascinante ensayo sobre la importancia de la educación en el contexto del mundo actual, en el que existen tantas amenazas a la racionalidad, en el escenario que el Dr. Leonardo Viniegra provocativamente llama «colapso civilizatorio». Además incluimos artículos originales sobre temáticas diversas: un estudio interesante sobre los pasantes de Servicio Social de la Facultad de Medicina de la UNAM; la producción científica de estudiantes de medicina en revistas médicas cubanas; las conductas de engaño en estudiantes de medicina en Argentina; las competencias genéricas de estudiantes de Obstetricia y Puericultura en Chile; las actitudes de médicos residentes sobre la especialidad de Medicina Preventiva y Salud Pública en España; y la evaluación formativa en la Facultad de Medicina de la UNAM, vista a través del lente del modelo Weisbord. Como podrán constatar los lectores, los artículos originales son de 5 países hispanoamericanos diferentes de Europa, Centro y Suramérica, lo que habla del progresivo interés internacional por publicar en nuestra revista, así como de la pluralidad de temas y objetos de estudio de los profesionales de la salud involucrados en educación.

Referencias

1. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing. *Standards for educational and psychological testing*. Washington, DC: AERA. 2014.
2. Downing SM. *Validity: on the meaningful interpretation of assessment data*. *Med Educ*. 2003;37:830-7.
3. Cook DA, Brydges R, Ginsburg S, Hatal R. *A contemporary approach to validity arguments: A practical guide to Kane's framework*. *Med Educ*. 2015;49:560-75.

4. Kane MT. Validating the interpretations and uses of test scores. *J Educ Meas.* 2013;50:1–73.
5. Cook DA, Kuper A, Hatala R, Ginsburg S. When assessment data are words: Validity evidence for qualitative educational assessments. *Acad Med.* 2016 Apr 5. [Publicación electrónica] doi: 10.1097/ACM. 0000000000001175.

Melchor Sánchez-Mendiola ^{a,b}

^a *Editor Investigación en Educación Médica*

^b *Coordinador de Desarrollo Educativo e Innovación Curricular, Ciudad Universitaria, Coyoacán, Ciudad de México, México*

Correos electrónicos: melchorsm@unam.mx,
melchorsm@gmail.com